

# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/JP05/006317

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: JP  
Number: 2004-107139  
Filing date: 31 March 2004 (31.03.2004)

Date of receipt at the International Bureau: 20 May 2005 (20.05.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 2 0 0 4 年 3 月 3 1 日

出 願 番 号  
Application Number: 特 願 2 0 0 4 - 1 0 7 1 3 9

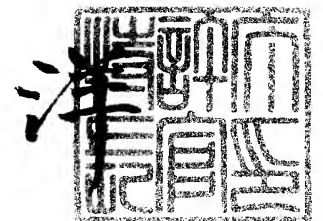
パリ条約による外国への出願  
に用いる優先権の主張の基礎  
となる出願の国コードと出願  
番号  
J P 2 0 0 4 - 1 0 7 1 3 9  
The country code and number  
of your priority application,  
to be used for filing abroad  
under the Paris Convention, is

出 願 人  
Applicant(s): 松下電器産業株式会社

2 0 0 5 年 4 月 2 7 日

特許庁長官  
Commissioner,  
Japan Patent Office

小 川



【書類名】 特許願  
【整理番号】 7048050012  
【提出日】 平成16年 3月31日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 7/00  
【発明者】  
    【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内  
    【氏名】 野本 昌子  
【特許出願人】  
    【識別番号】 000005821  
    【氏名又は名称】 松下電器産業株式会社  
【代理人】  
    【識別番号】 100082692  
    【弁理士】  
    【氏名又は名称】 蔵合 正博  
    【電話番号】 03-5210-2681  
    【ファクシミリ番号】 03-5210-2520  
【選任した代理人】  
    【識別番号】 100081514  
    【弁理士】  
    【氏名又は名称】 酒井 一  
    【電話番号】 03-5210-2681  
    【ファクシミリ番号】 03-5210-2520  
【手数料の表示】  
    【予納台帳番号】 013549  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1  
    【包括委任状番号】 0016258

【書類名】 特許請求の範囲

【請求項 1】

テキストから事実や意見などの情報をメタデータとして抽出するメタデータ抽出手段と、

メタデータ間の関連性を照合するメタデータ照合手段と、

関連するメタデータを統合して統合メタデータを生成するメタデータ統合手段とを備え、

前記メタデータ抽出手段は、事物に関する事実や意見などの情報を、事物のどのような点に着目して、あるいはどのような見地から述べられているかという「観点」と、その観点から具体的にどのような表現で記述されているかという「記述」の組からなる要素メタデータとして抽出する、観点・記述抽出手段をさらに有する、ことを特徴とする情報抽出システム。

【請求項 2】

前記メタデータ抽出手段は、テキスト中の事物名および事物に関して記述された情報に一定の規則にしたがって意味的属性を付与する意味属性付与手段をさらに有し、

前記観点・記述抽出手段が、事物に関して記述された情報に付与された意味的属性をも考慮して、少なくとも観点と記述の組からなる要素メタデータを抽出することを特徴とする請求項 1 に記載の情報抽出システム。

【請求項 3】

前記メタデータ抽出手段は、前記意味属性付与手段が意味属性を付与するために参照する規則を格納した意味属性付与規則記憶手段と、意味属性を付与したテキストを格納する意味属性付きテキスト記憶手段とをさらに有し、

前記観点・記述抽出手段が、前記意味属性付きテキスト記憶手段から意味属性が付与されたテキストを読み出して要素メタデータを抽出することを特徴とする請求項 2 に記載の情報抽出システム。

【請求項 4】

要素メタデータの話題の事物を推定する話題事物推定手段をさらに有し、前記メタデータ照合手段が、話題事物をも考慮してメタデータの照合を行う、ことを特徴とする請求項 2 または 3 に記載の情報抽出システム。

【請求項 5】

入力されるデータの種類として、テキスト、テキストの書誌事項の情報であるソース情報、テキストの作者に関する情報であるユーザ情報、またはこれらの組み合わせを判別する入力データ判定手段を有し、

前記メタデータ抽出手段が、ソース情報から観点と記述を抽出するためのソース観点・記述抽出規則を格納したソース観点・記述抽出規則記憶手段と、ユーザ情報から観点と記述を抽出するためのユーザ観点・記述抽出規則を格納したユーザ観点・記述抽出規則記憶手段と、のうち少なくともいずれか一方をさらに有し、

前記観点・記述抽出手段が、

前記ソース観点・記述抽出規則を参照してソース情報から少なくとも観点と記述の組からなるソースメタデータを抽出するか、または、

前記ユーザ観点・記述抽出規則を参照してユーザ情報から少なくとも観点と記述の組からなるユーザメタデータを抽出するか、のうち少なくとも一方をさらに実行して、テキストの内容だけでなく、テキストの書誌事項や作者に関する情報からもメタデータを抽出することを特徴とする請求項 1 乃至 4 のいずれかに記載の情報抽出システム。

【請求項 6】

前記メタデータ抽出手段が、さらにソース情報に意味的属性を付与するために意味属性付与手段が参照する規則を格納したソース意味属性付与規則記憶手段と、意味属性を付与したソース情報を格納する意味属性付きソース情報記憶手段を有し、前記ソース観点・記述抽出規則記憶手段が、意味属性付きソース情報から観点と記述を抽出するためのソース観点・記述抽出規則を格納するか、或いは

テキストの作者に関する情報であるユーザ情報に意味的属性を付与するために意味属性付与手段が参照する規則を格納したユーザ意味属性付与規則記憶手段と、意味属性を付与したユーザ情報を格納する意味属性付きユーザ情報記憶手段を有し、前記ユーザ観点・記述抽出規則記憶手段が、意味属性付きユーザ情報から観点と記述を抽出するためのユーザ観点・記述抽出規則を格納するか、の少なくともいずれか一方の組み合わせを有し、

前記観点・記述抽出手段は、

前記ソース観点・記述抽出規則を参照して意味属性付きソース情報から少なくとも観点と記述の組からなるソースメタデータを抽出するか、または、

前記ユーザ観点・記述抽出規則を参照して意味属性付きユーザ情報から少なくとも観点と記述からなるユーザメタデータを抽出するか、のうち少なくとも一方を実行することを特徴とする請求項 5 に記載の情報抽出システム。

#### 【請求項 7】

前記メタデータ照合手段が、要素メタデータの客観性と信頼性を判定する客観性・信頼性判定手段をさらに有することを特徴とする、請求項 1 乃至 6 のいずれかに記載の情報抽出システム。

#### 【請求項 8】

前記メタデータ照合手段が、要素メタデータ、ユーザメタデータ、ソースメタデータのうち少なくとも 1 種類を用いて、前記客観性・信頼性判定手段が要素メタデータの客観性・信頼性を判定するための客観性・信頼性判定規則を格納する客観性・信頼性判定規則記憶手段を有し、

前記客観性・信頼性判定手段が前記客観性・信頼性判定規則を参照して客観性・信頼性を判定することを特徴とする請求項 7 に記載の情報抽出システム。

#### 【請求項 9】

メタデータを表形式に整理してメタデータテーブルを生成するメタデータテーブル生成手段と、生成されたメタデータテーブルをユーザに提示するメタデータテーブル出力手段と、をさらに有することを特徴とする請求項 1 乃至 8 のいずれかに記載の情報抽出システム。

#### 【請求項 10】

ユーザからの要求を処理するユーザ要求処理手段をさらに有し、

前記メタデータテーブル生成手段が、ユーザ要求処理手段を通じて入力されたユーザ要求に合致するメタデータを用いてメタデータテーブルを生成することを特徴とする請求項 9 に記載の情報抽出システム。

【書類名】 明細書

【発明の名称】 情報抽出システム

【技術分野】

【0001】

本発明は、事物に関して記述された事実や意見等の内容を抽出する情報抽出システムに関する。

【背景技術】

【0002】

従来の情報抽出システムとしては、キーワードを抽出するもの、固有名や数値表現等を抽出するもの、5W1H等の事実に関する情報を抽出するもの、意見や評判を抽出するもの等が知られている。狭義の情報抽出は、文献『自然言語処理』（長尾他pp.438-441,1996）に紹介されているように、テキストの中心的な情報を抽出するもので、特定の分野のテキストを対象に抽出すべき情報のテンプレート（またはフレーム）を用意しておき、該当する情報を抽出するのがその典型である。一方、近年はテキスト中の事実だけでなく、意見や評判をも抽出しようとする研究が行われている。例えば、特許文献1は利用者が指定した物に関する意見を文書集合中から抽出するものである。

【特許文献1】 特開2003-203136

【発明の開示】

【発明が解決しようとする課題】

【0003】

しかしながら、特許文献1のような従来の意見情報抽出システムでは次のような不具合がある。

- (1) 意見や評判に関する表現としてあらかじめ辞書に登録された語しか抽出できない。
- (2) 抽出された意見がどのような事実と対応するものであるか確認する手段がない。
- (3) 意見の抽出元のテキストの書誌事項やテキストの作者に関する情報がえられず、原テキストやその作者に関する信頼性や記述の客観性を確認する手段がない。
- (4) 対象物を利用者が指定した物以外に拡張して意見を抽出するためには、類似した対象物名をグループとして登録しておかなければならない。
- (5) 利用者が指定した対象物に関する意見が書かれたテキストに別の物に関する記述が含まれていた場合、後者を排除する手段がない。

【0004】

また、前記の情報抽出システムのように特定の分野のみを対象として、5W1H的な情報を抽出するだけでは、一般のテキストからの事実の抽出は難しい、という問題がある。

【0005】

本発明は、かかる点に鑑みてなされたものであり、テキスト中の事物に関する事実や自由に記述された意見などの情報を精度良く、かつ、整理して抽出し、さらに客観性や信頼性を確認できるようにすることを目的とする。

【課題を解決するための手段】

【0006】

上記課題を解決するため、本発明の情報抽出システムは、テキストから事実や意見などの情報をメタデータとして抽出するメタデータ抽出手段、メタデータ間の類似度の比較を行うメタデータ照合手段、話題事物を推定するための話題事物推定手段、メタデータの照合結果に応じてメタデータを統合するメタデータ統合手段、統合したメタデータをもとに表形式のメタデータテーブルを作成するメタデータテーブル生成手段、メタデータテーブルをユーザに提示するメタデータテーブル出力手段、メタデータテーブルにユーザの要求を反映させるためにユーザの要求を受けつけるユーザ要求処理手段、テキストの書誌事項や作者に関する情報を利用するための入力データ判定手段、記述内容の客観性・信頼性を判定するための客観性・信頼性判定手段を有し、さらに前記メタデータ抽出手段が、テキスト中の事物名および事物に関して記述された情報に意味的属性を付与する意味属性付与手段、テキスト中の事物に関する事実や意見などの情報を少なくとも観点と記述の組から

なる要素メタデータとして抽出する観点・記述抽出手段、を有する構成をとる。

#### 【０００７】

この構成によれば、テキスト中の事物に関する事実や自由に記述された意見などの情報を意味的属性をも考慮して観点・記述として整理して抽出し、その話題となっている事物を推定することにより、これらの観点・記述およびその話題の事物をも考慮して、記述内容の対応関係や類似度を詳細に比較・分析することができ、また、テキストの書誌情報や作者に関する情報を用いることにより、客観性や信頼性の多面的な評価が可能になり、また、ユーザ要求を反映して抽出した内容を表形式に整理して提示することができる。

#### 【発明の効果】

#### 【０００８】

以上説明したように、本発明の情報抽出システムは、事実や意見などの情報をメタデータとして抽出するメタデータ抽出手段、メタデータ間の類似度の比較を行うメタデータ照合手段、話題事物を推定するための話題事物推定手段、メタデータの照合結果に応じてメタデータを統合するメタデータ統合手段、統合したメタデータをもとに表形式のメタデータテーブルを作成するメタデータテーブル生成手段、メタデータテーブルをユーザに提示するメタデータテーブル出力手段、メタデータテーブルにユーザの要求を反映させるためにユーザの要求を受けつけるユーザ要求処理手段、テキストの書誌情報や作者に関する情報を利用するための入力データ判定手段、記述内容の客観性・信頼性を判定するための客観性・信頼性判定手段を有し、さらに前記メタデータ抽出手段が、テキスト中の事物名および事物に関して記述された情報に意味的属性を付与する意味属性付与手段、テキスト中の事物に関する事実や意見などの情報を少なくとも観点と記述の組からなる要素メタデータとして抽出する観点・記述抽出手段、を有する構成により、

テキスト中の事物に関する事実や自由に記述された意見などの情報を意味的属性をも考慮して観点・記述として整理して抽出し、その話題となっている事物を推定することにより、これらの観点・記述およびその話題の事物をも考慮して、記述内容の対応関係や類似度を詳細に比較・分析することができ、また、テキストの書誌情報や作者に関する情報を用いることにより、客観性や信頼性の多面的な評価が可能になり、また、ユーザ要求を反映して抽出した内容を表形式に整理して提示することができる、という効果を有する。

#### 【発明を実施するための最良の形態】

#### 【０００９】

以下、本発明の実施の形態を、図面を参照して詳細に説明する。

#### 【００１０】

##### （実施の形態１）

図１は、実施の形態１に係る情報抽出システムの構成を示すブロック図である。この情報抽出システム１００は、入力されたテキスト中から事物に関する記述内容を、観点や記述の類似度を考慮して整理して抽出するためのシステムであって、テキストが入力される入力部１０２と、入力されたテキストから、事実や意見などの情報をメタデータとして抽出するメタデータ抽出部１０４と、抽出されたメタデータを照合するメタデータ照合部１０６と、メタデータ照合部１０６における照合結果に基づきメタデータを統合するメタデータ統合部１０８と、統合メタデータを格納する統合メタデータ記憶部１１０とを有する。さらにメタデータ抽出部１０４は、事物に関する事実や意見などの情報を、事物のどのような点に着目して、あるいはどのような見地から述べられているかという「観点」と、その観点から具体的にどのような表現で記述されているかという「記述」の組からなる要素メタデータとして抽出する観点・記述抽出部１２０と、要素メタデータを抽出するための規則を格納する観点・記述抽出規則記憶部１２２を有する。

#### 【００１１】

なお、情報抽出システム１００のハードウェア構成は、任意であって、特に限定されない。たとえば、情報抽出システム１００は、ＣＰＵや記憶装置（ＲＯＭ、ＲＡＭ、ハードディスクその他各種記憶媒体）を備えたコンピュータによって実現される。このように情報抽出システム１００がコンピュータによって実現される場合、情報抽出システム１００

は、この情報抽出システム 100 の動作を記述したプログラムを CPU が実行することによって所定の動作を行う。

#### 【0012】

この情報抽出システム 100 では、まず、入力部 102 で入力されたテキストを受け取る。そして、メタデータ抽出部 104 の観点・記述抽出部 120 が、事物に関して記述された内容を観点と記述の組からなる要素メタデータとして、観点、記述の組とテキストを識別するための要素メタデータ ID とともに抽出する。そして、メタデータ照合部 106 が、抽出された要素メタデータの観点、記述の類似度を比較・照合する。さらにメタデータ統合部 108 が、メタデータ照合部 106 の要素メタデータ間の類似度の判定結果にもとづいて、抽出された要素メタデータのうち類似の観点や記述をもつ要素メタデータをまとめて統合メタデータとし、統合メタデータ記憶部 110 に格納する。

#### 【0013】

ここで、メタデータとは、一般にコンテンツの内容や書誌事項等のコンテンツに関する情報を示すデータのことである。本発明では、テキスト中に記述された事物に関する事実や意見などの情報を、観点と記述の組として構成したものをコンテンツの内容に関するメタデータの基本単位とみなし、この基本単位に関する、少なくとも観点と記述を含む情報を特に要素メタデータとよぶ。「観点」とは、事物に関する事実や意見が、事物のどのような点に着目して、あるいはどのような見地から述べられているかということの意味し、「記述」とは、上記観点から具体的にどのような表現で記述されているかということの意味する。要素メタデータは、少なくとも、観点と記述の組から構成され、要素メタデータ ID を付与されることとする。要素メタデータ ID とは、要素メタデータの出現したテキストと、個々の要素メタデータを識別するために個々の要素メタデータに付与される要素メタデータの識別情報である。ただし、要素メタデータの観点は表層的な情報から得られない場合もあるので必須ではない。なお、1つの観点に対して複数の記述がある場合は、1つの観点に対して複数の記述を抽出してもよい。また、複数の要素メタデータを統合したものを統合メタデータとよぶ。

#### 【0014】

次いで、上記構成を有する情報抽出システム 100 について、具体例を用いてより詳細に説明する。図 2 は、入力されたテキストから事物に関して記述された内容を要素メタデータとして抽出するまでの一連の処理の概要を示す説明図である。

#### 【0015】

まず、メタデータ抽出部 104 内の観点・記述抽出部 120 が観点・記述抽出規則記憶部 122 に記憶された観点・記述抽出規則を参照し、入力部から入力されたテキスト内に規則に該当する観点・記述の組があるかどうか調べる。観点・記述抽出規則と規則の構成要素定義の例を図 3 に示す。なお、これ以降規則で同様の構成要素を用いる場合は定義を省略する。各規則には観点・記述を抽出するためのボタンと、ボタン中で観点、記述に該当する箇所が示されている。図 3 の例では、観点や記述に該当する箇所は ” ( ) ” でマークされており、マークされた部分は先頭から順に \$1、\$2、… と参照される。例えば、規則 1 の場合は、<「は」><漢字／平仮名連続><「が」または「も」>、<漢字／平仮名連続>、<形容詞語尾 1> がこの順でテキスト中に出現した場合、規則のボタンと一致する。テキスト中のこのボタンが現れる部分で、ボタン中、最初の ” ( ) ” で括られた <漢字／平仮名連続> に相当する部分が \$1 として参照され、2 番目の ” ( ) ” で括られた <漢字／平仮名連続><形容詞語尾 1> に相当する部分が \$2 として参照される。規則にしたがって、前者は観点、後者は記述として抽出される。

#### 【0016】

図 3 の規則 1 を図 2 のテキスト 1 に適用した場合、1 文目の ” 開口部 ” が観点、 ” かなり大きい ” が記述に該当する。図 2 の観点・記述認定例はテキスト内の観点・記述対に、1 から順に番号を与え、観点を表現の始まりと終わりを < VIEW・( 観点・記述対の番号 ) > … < /VIEW ( 観点・記述対の番号 ) >、記述の表現の始まりと終わりを < DESC ( 観点・記述対の番号 ) > … < /DESC ( 観点・記述対の番号 ) > でマークしたものである。なお



、観点・記述対に対する番号の与え方はこれに限定されるものではない。例えば、テキストの識別情報とテキスト内での観点・記述対の番号を組み合わせただけのものでもよい。

#### 【0017】

なお、例えば「容量が20リットルと大きい」のように1つの観点（この例では「容量」）に対して、「20リットル」「大きい」と複数の記述がある場合には、これらを同じ観点に対する異なる2つの記述として抽出してもよい。本発明の観点と記述の抽出規則例では、同じ観点に対して複数の異なる記述を抽出する場合、これらの記述を記号「||」を用いて例えば「\$1||\$2」（ただし、\$1、\$2 は記述）のように示す。

#### 【0018】

一方、例えば「容量が旅行用には小さい」のように1つの観点（この例では「容量」）に対して、用途が「旅行用」に限定された場合は「小さい」というように、記述間に限定的な関係がある場合は、複数の記述（この例では「旅行用」と「小さい」）をまとめて1つの記述としてあつかってもよい。本発明の観点と記述の抽出規則例では、同じ観点に対して複数の記述に関連した記述としてまとめて抽出する場合、これらの記述を記号「&&」を用いて例えば「\$1&&\$2」（ただし、\$1、\$2 は記述）のように示す。

#### 【0019】

次に観点・記述抽出部120は、上記の観点・記述抽出規則に該当すると認定された観点・記述の組に、観点・記述対が出現したテキストと個々の観点・記述対を識別するための要素メタデータIDを付与し、規則にしたがって抽出する。抽出例を図2の要素メタデータ抽出結果の表に示す。この抽出結果表において、要素メタデータIDの最上段に記載された「1-1」のうち、左側の「1」は、この観点「開口部」・記述「かなり大きい」がテキスト1から抽出されたものであることを示し、右側の「1」は、観点「開口部」・記述「かなり大きい」がテキスト1を検索したときに第1番目（つまり最初）にヒットした観点・記述であることを示す。なお、本実施例では要素メタデータIDを<テキストID>-<観点・記述対のテキスト内での番号>という型式で付与することとしたが、要素メタデータIDの型式は、テキストの識別と観点・記述対の識別が可能なものであれば、これに限定されるものではない。

#### 【0020】

続いて、抽出された要素メタデータをメタデータ照合部106が観点・記述の類似度により比較・照合する。観点・記述の類似度の照合方法としては、特に限定されないが、たとえば、観点または記述の構成語の概念的な類似性をシソーラス、類義語辞書等を用いて比較する方法、観点または記述の構成語の構文的な関係の類似度を推定する方法などを用いることができ、本発明はいずれかの方法に限定されるものではない。ここでは仮に、観点や記述から以下のように助詞や語尾を除く構成語と、構成語と構成語との間の構文的関係とを取り出して照合することとする。図2の観点から取り出される構成語と構成語間の構文的関係は以下になる。

#### 【0021】

開口部 →（構成語）：開口、部（構文的関係）連体修飾

ファスナーの開閉 →（構成語）：ファスナー、開閉（構文的関係）連体修飾

皮の感触 →（構成語）：皮、感触（構文的関係）連体修飾

皮の手触り →（構成語）：皮、手触り（構文的関係）連体修飾

色合い →（構成語）：色合い

次に観点「皮の感触」、「皮の手触り」の構成語のうち「感触」「手触り」をシソーラスにより同義語と認定し、他の構成語「皮」および構文的関係も一致していることから、2つの観点「皮の感触」「皮の手触り」は同義と判定することとする。また、記述についても同様にして、同義の記述を求めることとすると、要素メタデータID1-3と2-2から「しっとりやさしい」という同義の記述がえられる。

#### 【0022】

次に、要素メタデータ間の類似度の照合結果にもとづいて、メタデータ統合部108が、要素メタデータを統合し、統合メタデータとして統合メタデータ記憶部110に格納す

る。メタデータの統合の仕方は、特に限定されないが、ここでは、

同義の観点をもつメタデータを統一する

同義の観点をもつメタデータで同義の記述があれば統一する

こととする。図2の例では、観点のうち「皮の感触」「皮の手触り」が同義と判定されたので、これらの観点を統合することとする。また、これらの観点と対になっている記述「しっとりやさしい」と「驚くほどなめらかだ」は同義とはみなされないので、統合しない。このようにして統合処理を行なった後のメタデータの例を図4に示す。

#### 【0023】

このように本実施の形態によれば、テキスト中の事物に関する記述内容を、事物の観点や記述の類似度を考慮して整理して抽出することができる。

#### 【0024】

(実施の形態2)

図5は本発明の実施の形態2に係る情報抽出システムの構成を示すブロック図である。この情報抽出システム200は、図1に示す実施の形態1に対応する情報抽出システム100と同様の基本的構成を有しており、同一の構成要素には同一の符号を付し、その説明を省略する。

#### 【0025】

本実施の形態の特徴は、前記メタデータ抽出部104が、テキストに意味的属性を付与する意味属性付与部202、意味属性付与規則記憶部204、意味属性付きテキスト記憶部206をさらに有することである。意味属性付与部202の処理結果、つまり意味属性が付与されたテキスト(意味属性付きテキスト)は、意味属性付きテキスト記憶部206に格納される。この場合、観点・記述抽出部120は意味属性付きテキスト記憶部206に格納された意味属性付きテキストに対して抽出を行う。

#### 【0026】

意味属性付与部202はテキスト中の事物名、数値関連表現(時、数量、金額等)等を認定し、これらに意味的属性を付与する。事物名や数量表現に意味的属性を付与する方法としては、特に限定されないが、たとえば、キーワード毎にその意味属性を記載した辞書を用いる方法や、文献「福本他: "固有名詞抽出における日本語と英語の比較"、情報処理学会研究会報告 98-NL-126, pp.107-114, 1998」に示される固有名詞抽出技術を利用する方法などを用いることができる。

#### 【0027】

ここで意味属性とは、たとえば、事物名や数量表現を各表現の意味により分類した意味分類である。意味属性が詳細度のレベルをもつ場合や、該当の表現が一般的な表現の別表現であり、正規化された形を示す必要がある場合は、詳細度レベルや正規化された表現を意味属性の詳細情報として併記してもよい。

#### 【0028】

以下では、意味属性抽出規則を用いて事物名と数量表現に意味属性を付与する例を説明する。

#### 【0029】

まず、意味属性付与部202が意味属性付与規則記憶部204に格納された意味属性付与規則を参照して、入力部から入力されたテキスト内に規則に該当する意味属性をもつ表現があるかどうか調べ、テキスト中の該当する表現と意味属性をマークし、意味属性付きテキストとして意味属性付きテキスト記憶部206に格納する。図6に、入力されたテキストと意味属性を付与されたテキストの例、図7に意味属性付与規則の例と意味属性付与規則構成要素定義の例を示す。なお、意味分類付与規則構成要素の定義方法は例えば構成要素毎に分割した語リストのファイルとするなど、複数ファイルで用意してもよい。図7の規則例には、テキスト中で該当する意味属性をもつ表現を検出するためのパターン、各パターンに合致する表現に付与される意味属性および詳細情報が示されている。この例では詳細情報のうちvalは数値表現の正規化された値を示し、unitは数量単位の表現の正規化形であり、typeは意味属性の下位分類を示すこととする。

### 【0030】

図7の規則を図6のテキスト1に適用した場合、規則1により「20リットル」の意味属性はQUANT(数量)、詳細情報は[unit=1(単位:1),val=20(数値:20)]と認識され、規則2により「容量」がQUANT—TYPE(数量分類)、また規則3により「A社」がORGANIZATION(組織名)、詳細情報は[type=company(会社名)]、等と認識され、各々該当する意味属性と詳細情報を付与され、図6に示したような意味属性付きテキストとして意味属性付きテキスト記憶部206に格納される。

### 【0031】

次に、観点・記述抽出部120が、意味属性付きテキスト記憶部206に格納された意味属性付きテキストから観点・記述の組を、必要ならばその他の意味的属性に関する情報とともに、要素メタデータとして抽出する。意味属性付きテキストの例と観点・記述認定例を図8、観点・記述の抽出のための規則の例と観点・記述抽出規則の構成要素の定義例を図9に示す。図9に示した観点・記述抽出規則と実施の形態1の図3に示した観点・記述抽出規則の違いは、図9では、テキストに付与された意味属性がパタンの一部として記述されうることである。例えば、図9の規則1では<QUANT—TYPE>,</QUANT—TYPE>で囲まれたタグ開始記号以外の任意文字列、即ち、QUANT—TYPE(数量分類)という意味属性を付与された文字列が観点であり、<QUANT>,</QUANT>で囲まれたタグ開始記号以外の任意文字列、即ち、QUANT(数量)という意味属性を付与された文字列が前記観点に対応する1つめの記述であることが示されている。図9の規則1を図8のテキスト1に適用した場合、QUANT—TYPEの意味属性を付与された「容量」が観点であり、QUANTの意味属性を付与された「20リットル」がこの観点に対応する1つめの記述である。図8の意味属性付きテキスト1,2に対して図9の規則を適用して抽出した観点と記述、およびそれらの意味属性および詳細情報を要素メタデータとして抽出した結果の例を図10に示す。

### 【0032】

次にメタデータ照合部106がメタデータ間の類似度を照合する。メタデータ照合部106の照合方法は実施の形態1と同様であるが、照合の際に要素メタデータのうち、意味属性や詳細情報を用いてもよい。ここでは実施の形態1と同様に、図10の要素メタデータから同義の観点や記述を求めることとすると、1-2と2-1の「製品分類」と「バッグ」、1-3と2-2の「製品名」と「A200」が各々同義の観点と記述と判定され、1-4aと1-4bと2-3の「容量」が同義の観点と判定される。

### 【0033】

次に、実施の形態1と同様に、メタデータ統合部108が要素メタデータを統合し、統合メタデータとして統合メタデータ記憶部110に格納する。ここでは仮に実施の形態1と同様の方法で統合され、統合メタデータ記憶部110に格納された統合メタデータの例を図11に示す。図11において、同義の観点と記述をもつ要素メタデータである1-2と2-1、1-3と2-2が各々統合されている。また、同じ観点「容量」で「20リットル」「大きい」「不十分だ」の3つの異なる表現の記述が統合されており、数量である「20リットル」がこの製品の容量として「大きい」「不十分だ」と評価されていることがわかる。

### 【0034】

このように本実施の形態によれば、テキスト中の事物に関する記述内容を、意味的分類をふまえた上で観点と記述としてより精密に抽出し、さらに詳細に比較、分析することができる。

### 【0035】

#### (実施の形態3)

図12は本発明の実施の形態3に係る情報抽出システムの構成を示すブロック図である。この情報抽出システム300は、図5に示す実施の形態2に対応する情報抽出システム200と同様の基本的構成を有しており、同一の構成要素には同一の符号を付し、その説明を省略する。

### 【0036】

本実施の形態の特徴は、ユーザからの要求を処理するユーザ要求処理部302と、メタデータを表形式に整理してメタデータテーブルを生成するメタデータテーブル生成部304と、メタデータテーブル生成部304が生成したメタデータテーブルをユーザに提示するメタデータテーブル出力部306と、話題事物推定部310と話題事物推定規則記憶部312をさらに有することである。

#### 【0037】

ここで、「話題事物」とは、各要素メタデータがどの事物について記述されているかという、要素メタデータの話題の事物名のことである。この話題事物は、事物名を表す要素メタデータのいずれかの記述から選択される。話題事物の候補となりうる事物名は、特に限定されないが、人名、地名、組織名、イベント名、生物や人工物の名およびそれらの分類（例：製品名、製品分類）等がある。

#### 【0038】

次いで、上記構成を有する情報抽出システム300について、具体例を用いてより詳細に説明する。今、以下のテキスト1，2があるとする。

テキスト1：バッグA200は容量が不十分だし、バッグA300は容量があまりに大きい。

テキスト2：バッグA200は容量が20リットルで、バッグA300の容量は30リットル。

前記テキストが入力部から入力され、意味分類が付与され、観点・記述が認定され、要素メタデータが抽出されるまでの処理の流れは実施の形態2と同様であり、説明を省略する。このテキストに対して意味分類を付与し、観点・記述を認定した結果の例と、要素メタデータの抽出結果の例を図13に示す。

#### 【0039】

次に話題事物推定部310は、話題事物推定規則記憶部312に格納された話題事物推定規則にしたがって、テキスト内の話題事物を推定する。話題事物の推定方法は話題事物推定規則を用いるものであれば、特に限定されない。話題事物推定部310が話題事物推定規則を用いて直接話題事物を推定することとしてもよいし、まず話題事物候補となる要素メタデータの種類を決定し、その後に話題事物推定規則を用いて推定することとしてもよい。入力されるテキストが例えば会社名と人名等、複数の種類の話題をもつ可能性がある場合は、複数の話題事物推定候補を想定し、前記話題事物推定部310が、適当な話題事物を選択できるようにしておくことがのぞましい。ここでは話題事物候補が「観点が製品名または人名」である要素メタデータの記述、と規定されている場合を例に説明する。今、観点が製品名または人名である要素メタデータの記述が話題事物の候補と規定されていることから、テキスト1，2とも、製品名を観点にもつ要素メタデータの記述である、

A200

A300

が話題事物候補となる。

#### 【0040】

次に話題事物推定部310は、話題事物推定規則記憶部312に格納された話題事物推定規則にしたがって、テキスト内の話題事物を推定する。ここでは仮に、条件部に記述したボタンとのマッチングにより話題事物推定をおこなうこととし、話題事物推定規則と話題事物推定規則構成要素定義の例を図14に示す。なお、図14の規則では、話題事物候補の観点・記述や意味分類によって規則を分けていないが、これらをボタン中に指定して、話題事物候補の種類によって異なる規則を適用することとしてもよい。図14の規則を用いて前記テキスト1，2から話題事物を推定した例を図15に示す。図15の要素メタデータID1-1、1-4、2-1、2-4の要素メタデータについては図14の規則3が適用され、図15のID1-2、1-5、2-2、2-5の要素メタデータについては図14の規則1が適用され、図15のID1-3、1-6、2-3、2-6のメタデータについては図14の規則2が適用されている。

#### 【0041】

続いてメタデータ照合部106が、メタデータを照合する。メタデータの照合方法は実施の形態1または2と基本的に同様であるが、話題事物の推定結果をも考慮して照合する。例えば同じ話題事物をもつ要素メタデータ（図15の例では要素メタデータID1-1, 1-2, 1-3, 2-1, 2-2, 2-3が同じ話題事物をもち、1-4, 1-5, 1-6, 2-4, 2-5, 2-6が同じ話題事物をもつ）を認定する。また、実施の形態1または2と同様にして、同義の観点や記述をもとめることとすると、同義の観点と記述をもつ要素メタデータは1-1と1-4と2-1と2-4, 1-2と2-2, 1-5と2-5であり、同義の観点をもつ要素メタデータは1-3と1-6と2-3と2-6がえられる。

#### 【0042】

次に、実施の形態1と同様にして、メタデータ統合部108が要素メタデータを統合し、統合メタデータとして統合メタデータ記憶部110に格納する。要素メタデータの統合の仕方は限定されないが、ここでは例として、

同じ話題をもつ事物を統合する

同じ話題で同義の観点をもつ要素メタデータを統一する

同じ話題で同義の観点をもつ要素メタデータで同義の記述があれば統一する

とした場合に、統合メタデータ記憶部110に格納された統合メタデータの例を図16に示す。図16において、同じ話題事物をもち、同義の観点と記述をもつ要素メタデータ1-1と2-1, 1-2と2-2, 1-4と2-4, 1-5と2-5は各々統合されている。また、同じ話題事物と同義の観点をもつ要素メタデータ1-3と2-3, 1-6と2-6は各々統合されており、「A200」の「容量」が「20リットル」で「不十分だ」と評価されている一方、「A300」の「容量」が「30リットル」で「あまりに大きい」と評価されていることがわかる。

#### 【0043】

メタデータテーブル生成部304は統合メタデータ記憶部110に格納された統合メタデータを参照して、メタデータテーブルを生成し、メタデータテーブル出力部306を通じてユーザに提示する。さらにユーザ要求処理部302を通じてユーザ要求が入力された場合は、メタデータテーブル生成部304は同様にしてユーザ要求を反映したメタデータテーブルを再度生成し、メタデータテーブル出力部306を通じてユーザに提示する。

#### 【0044】

ここでは、ユーザ要求の指定にしたがって、メタデータテーブルを生成する場合を説明する。まず、ユーザ要求処理部302を通じてユーザ要求が入力される。ユーザ要求処理部302に入力されるユーザ要求は、話題事物を含めた要素メタデータの一部、あるいはこれらの組み合わせのいずれかを指定するものとする。ユーザ要求の例としては、例えば

（話題事物：A200）

かつ

（観点：容量）

という条件を満たす要素メタデータが指定されたとする。ユーザ要求処理部302は指定されたユーザ要求の指定形式をチェックし、問題がなければユーザ要求をメタデータテーブル生成部304に送る。

#### 【0045】

なお、この例では、ユーザ要求が上記の形式で入力されることとしたが、ユーザ要求が自由なテキスト（例：「A200の容量が知りたい」）で入力された場合は、ユーザ要求処理部302がメタデータ抽出部104にユーザの入力したユーザ要求を送り、抽出された要素メタデータとそれらの構文的な関係から要素メタデータ間の関係を上記のような形に解析することとしてもよい。メタデータテーブル生成部304はユーザ要求の分析結果にしたがって、統合メタデータの中から該当する要素メタデータを選別し、メタデータテーブルを生成する。例えばユーザ要求の中に話題事物があれば、この話題事物を話題にもつ要素メタデータを統合メタデータの中から選別した後に、指定された観点をもつメタデータを選別し、それらを対象としたメタデータテーブルを生成し、メタデータテーブル出

力部 3 0 6 がメタデータテーブルを出力する。図 1 7 にユーザ要求にもとづくメタデータテーブルの例を示す。話題事物が「A 2 0 0」で観点が「容量」の要素メタデータのみがテーブルとして出力されている。

#### 【0 0 4 6】

このように本実施の形態によれば、テキスト中の事物に関する記述内容を、観点と記述として整理するだけでなく、さらに話題の事物をも推定することにより、抽出された内容をより正確に比較、分析することが可能となり、さらに、話題事物を含めた要素メタデータを、ユーザの指定にもとづいて整理したメタデータテーブルをユーザに提示することにより、ユーザが要求する情報を精度よく、かつわかりやすく整理して提示することができる。

#### 【0 0 4 7】

##### （実施の形態 4）

図 1 8 は本発明の実施の形態 4 に係る情報抽出システムの構成を示すブロック図である。この情報抽出システム 4 0 0 は、図 1 2 に示す実施の形態 3 に対応する情報抽出システム 3 0 0 と同様の基本的構成を有しており、同一の構成要素には同一の符号を付し、その説明を省略する。

#### 【0 0 4 8】

本実施の形態の特徴は、前記入力部がソース情報およびユーザ情報をも受けとり、入力データの種別を判定する入力データ判定部 4 0 2 と、入力データ判定規則を格納するための入力データ判定規則記憶部 4 0 4 をさらに有し、メタデータ抽出部 1 0 4 が、意味属性付与部 2 0 2 がソース情報に意味属性を付与するためのソース意味属性付与規則を格納するソース意味属性付与規則記憶部 4 0 6 と、意味属性を付与されたソース情報を格納するための意味属性付きソース情報記憶部 4 0 8 と、意味属性付与部 2 0 2 がユーザ情報に意味属性を付与するためのユーザ意味属性付与規則を格納するユーザ意味属性付与規則記憶部 4 1 6 と、意味属性を付与されたユーザ情報を格納するための意味属性付きユーザ情報記憶部 4 1 8 と、観点・記述抽出部 1 2 0 が意味属性付きソース情報から観点と記述を抽出するためのソース観点・記述抽出規則記憶部 4 1 0 と、観点・記述抽出部 1 2 0 が意味属性付きユーザ情報から観点と記述を抽出するためのユーザ観点・記述抽出規則記憶部 4 2 0 と、をさらに有し、メタデータ照合部 1 0 6 が客観性・信頼性判定部 4 1 2、客観性・信頼性判定規則記憶部 4 1 4 をさらに有することである。

#### 【0 0 4 9】

客観性・信頼性判定部 4 1 2 は少なくとも要素メタデータ、またはソース情報からえられたメタデータ、またはユーザ情報からえられたメタデータのいずれかをを用いて要素メタデータの客観性と信頼性を判定し、判定結果を要素メタデータの評価データとする。メタデータ統合部 4 1 4 は要素メタデータに加えて、ソース情報からえられたメタデータ、ユーザ情報からえられたメタデータ、要素メタデータの前記評価データを統合結果として抽出する。また、ユーザはユーザ要求処理部 3 0 2 から要素メタデータだけでなく、ユーザ情報またはソース情報からえられたメタデータをも用いて必要な情報を指定し、対応するメタデータテーブルをうることができる。

#### 【0 0 5 0】

ここで、ソース情報とは入力されるテキストに関する書誌事項の情報を指すものとする。ソース情報の例としてはテキストの種別、入手元、作成者分類、作成者、組織名、作成日時等がある。ソース情報の書式は特に限定されないが、テキストとの対応関係を示す、テキスト ID の情報とともに入力することとする。また、他の情報やテキストとともに入力される場合はファイル内の該当箇所の情報とともに入力されることがのぞましく、さらに、意味属性付与の要不要の情報とともに入力されるようにしてもよい。

#### 【0 0 5 1】

また、ユーザ情報とは入力されるテキストの作者に関する情報を指すものとする。ユーザ情報の例としては、ユーザの性別、年齢、職業、勤務先、趣味、等がある。ユーザ情報の書式は特に限定されないが、テキストとの対応関係を示すため、テキスト ID の情報と

ともに入力されることとする。また、他の情報やテキストとともに入力される場合はファイル内の該当箇所の情報とともに入力されることがのぞましく、さらに、意味属性付与の要不要の情報とともに入力されるようにしてもよい。また、ソース情報、ユーザ情報を観点と記述の組として構成したものを各々ソースメタデータ、ユーザメタデータとよぶ。ソースメタデータおよびユーザメタデータには対応するテキストと個々のソースメタデータまたはユーザメタデータを識別するためのソースメタデータIDまたはユーザメタデータIDが付与される。ソースメタデータIDおよびユーザメタデータIDの書式は特に限定されないが、テキストとの対応関係をとる必要があるため、対応するテキストIDが推定可能な書式とすることがのぞましい。

#### 【0052】

次いで、上記構成を有する情報抽出システム400について、具体例を用いてより詳細に説明する。まず、入力データ判定部402が入力部102から入力されたファイルのデータの種類、複数種類のデータの場合は、ソース情報またはユーザ情報の該当箇所、意味属性付与処理の要・不要を入力データ判定規則記憶部404に格納された入力データ判定規則を用いて判定する。データの種類とは、ファイル内のデータの種類を特定するための情報で、データがテキストのみか、ソース情報のみか、ユーザ情報のみか、ソース情報、ユーザ情報、テキストの少なくとも2種類以上を含むか等の情報である。また、ソース情報またはユーザ情報の該当箇所とはファイル内でこれらの情報を記載した部分を特定するための情報であり、ファイル内に複数種類の情報がない場合や、システム内でデータの種類により規定された値を用いる場合等は不要である。

#### 【0053】

また、意味属性付与処理の要・不要とはユーザ情報やソース情報があらかじめ観点と記述またはそれに対応する形に整理して記述されており、意味属性付与処理が不要な場合に、意味属性付与部202の意味属性付与処理を経由せずに直接、観点・記述抽出部120が観点・記述抽出処理をおこなえるようにするための情報である。無条件に意味属性付与処理をおこなう場合や、システム内でデータの種類により意味属性付与処理をおこなうかどうか規定されている場合等は、この情報は不要である。

#### 【0054】

図19に入力データ判定規則の例を示す。例えばファイル名がA0001.BIBの場合は、規則2にしたがって、入力されたデータは規則2にしたがって、ソース情報のみで、対応するテキストは“A0001”と判定される。なお、入力データの種類や対応するテキストIDを判定するための条件は、ファイル名以外のものを用いてもかまわない。また、入力データの種類として、図19の規則では規則4で2種類以上のデータを含むかという判定をしているが、データの種類の組み合わせによってさらに入力データの種類を細分してもかまわない。また、図19の規則ではソース情報やユーザ情報の該当箇所をファイル内の行数で特定する例を示したが、この方法に限定されるものではなく、例えば、バイト位置等を用いて特定してもかまわない。

#### 【0055】

次に、意味属性付与部202が、意味属性付与規則記憶部204に格納された意味属性付与規則、またはソース意味属性付与規則記憶部406に格納されたソース意味属性付与規則、またはユーザ意味属性付与規則記憶部416に格納されたユーザ意味属性付与規則を用いて、テキストまたはソース情報またはユーザ情報に意味属性を付与し、意味属性付きテキストまたは意味属性付きソース情報または意味属性付きユーザ情報として意味属性付きテキスト記憶部206または意味属性付きソース情報記憶部408または意味属性付きユーザ情報記憶部418に格納する。図20にテキストのソース情報、ユーザ情報、意味属性付きソース情報、意味属性付きユーザ情報の例、図21にソース意味属性付与規則、ユーザ意味属性付与規則の例を示す。意味属性付与規則を用いたテキストの意味属性の付与情報、ソース意味属性付与規則を用いた意味属性の付与方法、ユーザ意味属性付与規則を用いた意味属性の付与方法は、実施の形態2と同様であり、説明を省略する。なお、ユーザ情報やソース情報の意味属性付与処理が不要と指定されていた場合は、意味属性付



与処理はおこなわない。

#### 【0056】

次に、観点・記述抽出部120が、観点・記述抽出規則記憶部122に格納された観点・記述抽出規則、またはソース観点・記述抽出規則記憶部410に格納されたソース観点・記述抽出規則、またはユーザ観点・記述抽出規則記憶部420に格納されたユーザ観点・記述抽出規則を用いて、意味属性付きテキスト記憶部206、または意味属性付きソース情報記憶部408、または意味属性付きユーザ情報記憶部418に格納された、意味属性付きテキストまたは意味属性付きソース情報または意味属性付きユーザ情報から、少なくとも観点と記述の組からなる要素メタデータ、またはソースメタデータ、またはユーザメタデータを抽出する。ソースメタデータまたはユーザメタデータを抽出する際、各ソースメタデータにはソースメタデータID、ユーザメタデータにはユーザメタデータIDが付与される。なお、本実施例ではソースメタデータIDおよびユーザメタデータIDを、各々、〈テキストID〉-S〈観点・記述対のソース情報内での番号〉、〈テキストID〉-U〈観点・記述対のユーザ情報内での番号〉という形式で付与することとしたが、ソースメタデータIDの形式およびユーザメタデータIDの形式は、これに限定されるものではない。

#### 【0057】

図22にソース観点・記述抽出規則およびユーザ観点・記述抽出規則の例を示す。また、図23に図20の意味属性付きソース情報および意味属性付きユーザ情報から図22のソース観点・記述抽出規則またはユーザ観点・記述抽出規則を用いて抽出したソースメタデータ抽出結果およびユーザメタデータ抽出結果の例を示す。観点・記述の抽出方法は、ソース観点・記述抽出規則またはユーザ観点・記述抽出規則を用いる場合も、実施の形態2の要素メタデータの場合と同様であり、説明を省略する。

#### 【0058】

図25にテキスト例と実施の形態2または3と同様にして意味属性を付与したテキストの例、図26に観点・記述抽出規則の例を示す。また、図25の意味属性付きテキストから、図26の観点・記述抽出規則を用いて、実施の形態2または3と同様にして観点・記述を抽出し、さらに図14の話題事物推定規則を用いて実施の形態3と同様にして話題事物を推定した場合の、要素メタデータの抽出結果の例を図27に示す。なお、要素メタデータは一部のみを示している。また、実施の形態3では、話題事物の推定に関して、テキストからえられる情報だけを用いて推定する方法を説明したが、他にソース情報やユーザ情報からえられるメタデータを用いてもよい。

#### 【0059】

次に、客観性・信頼性判定部412が、テキストから抽出された要素メタデータとソースメタデータとユーザメタデータのうち、少なくとも1・種類以上を用いて、客観性・信頼性判定規則記憶部414に格納された客観性・信頼性判定規則にしたがって要素メタデータの評価として、メタデータの客観性および信頼性を判定する。ここで、要素メタデータの客観性とは、要素メタデータが客観的に記述されているかどうかを示し、例えば事実として記述されているならば客観性は高く、意見として記述されているならば客観性は低いと考えられる。客観性は数値として表現してもよいし、ある閾値または判定条件により「事実」「意見」等の分類で表現してもよい。

#### 【0060】

また、要素メタデータの信頼性とは、要素メタデータが信頼できるかどうかを示し、例えば個人のホームページに意見として書かれた記述の信頼性は比較的低く、新聞記事に事実として書かれた記述の信頼性は高い、等と考えられる。要素メタデータの客観性・信頼性の判定には少なくとも要素メタデータ、ソースメタデータ、ユーザメタデータのいずれか一つを用いることとするが、これらの他に構文的な情報、統計的な情報等を合わせて用いてもかまわない。図24に客観性・信頼性判定規則例を示す。

#### 【0061】

今、前記と同様にして、テキスト1、テキスト2、テキスト3、テキスト4に対応して



それぞれ以下のようなソースメタデータとユーザメタデータが抽出されたとする。

【0062】

テキスト1

ソースメタデータ

観点： テキストの分類

記述：会社のweb ページ

テキスト2

ソースメタデータ

観点：テキストの分類

記述：個人のweb ページ

ユーザメタデータ

観点：性別

記述：男性

テキスト3

ソースメタデータ

観点：テキストの分類

記述：個人のweb ページ

ユーザメタデータ

観点：性別

記述：女性

テキスト4

ソースメタデータ

観点：テキストの分類

記述：個人のweb ページ

ユーザメタデータ

観点：性別

記述：男性

上記のソースメタデータおよびユーザメタデータと図27の要素メタデータの抽出結果に対して、図24の客観性・信頼性判定規則を用いて抽出した客観性・信頼性判定結果例を図28に示す。

【0063】

また、図24の客観性・信頼性判定規則ではソースメタデータの観点「テキストの分類」と対応する記述を規則の条件の一部に用いたが、他の観点と記述の組を用いてもよい。例えば「作成日」を用いて作成日が古い要素メタデータは信頼性が低いと判定する、あるいは「作成者」を用いて特定の人の書いたテキストの信頼度を上げる、または下げるといようにしてもよい。また、要素メタデータと他の情報を組み合わせる場合、例えば、統計的な情報と組み合わせ、同じ観点に対して多数の類似の内容の記述をもつ要素メタデータの信頼度を上げる、あるいは、多数の人の記述と異なる内容の記述をもつ要素メタデータの信頼度を下げる、というようにしてもよい。なお、図24の客観性・信頼性判定規則では、1規則で客観性と信頼性を同時に判定しているが、1規則でいずれか一方を判定するようにしてもかまわない。

【0064】

次にメタデータ照合部106は、要素メタデータの類似度を照合する。メタデータ照合部106の類似度の照合方法は特に限定されない。ここでは実施の形態1または2または3と同様とするが、客観性・信頼性のデータをも用いて類似度を比較してもよい。

【0065】

次にメタデータ統合部108が要素メタデータとソースメタデータとユーザメタデータと評価を含めた要素メタデータを統合し、統合結果を統合メタデータ記憶部110に格納する。統合の仕方は限定されないが、ここでは例として、

(要素メタデータ)

同じ話題をもつメタデータを統合する

同じ話題で同義の観点をもつメタデータを統合する

同じ話題で同義の観点をもつメタデータで同義の記述があれば統合する

同じ話題で同義の観点と同義の記述をもつメタデータで意味属性が同じなら統合する

とした場合に、統合メタデータ記憶部 110 に格納されたメタデータの統合結果を図 29 に示す。要素メタデータの一部は記述を省略している。この場合、客観性、信頼性ともに高い情報、すなわち事実である可能性が高い情報として、「A200」という「バッグ」の「容量」が「20リットル」であるという情報があり、また、それに関する客観性の低い情報、すなわち意見と思われる情報として、会社のホームページではその容量が「大きい」と評価されているが、個人のホームページでは、「海外出張用」には男性1名に「不十分だ」と評価され、「国内出張用」には女性1名に「あまりに大きい」と評価され、男性1名に「十分だ」と評価されていることが分かる。

#### 【0066】

次にメタデータテーブル生成部 304 がメタデータテーブルを生成し、メタデータテーブル出力部 306 を通じてユーザに提示し、ユーザからユーザ要求の指定があれば、再度生成したメタデータテーブルをユーザに提示するまでの流れは実施の形態3と同様である。ここでは、図 29 のメタデータの統合結果をもとに作成したメタデータテーブルを一度ユーザに提示した後、次のようなユーザ要求の指定を受けて、メタデータテーブル生成部 304 が再度メタデータテーブルを生成する場合を一例として説明する。

#### 【0067】

ユーザ要求の指定

話題事物：A200

客観性：0

テキストの種類：個人web ページ

この指定は、個人web ページのテキストに書かれた、「A200」という事物についての「客観性が0」の記述、すなわち意見を求めるものである。

#### 【0068】

上記のユーザ要求の指定により、実施の形態3と同様にして生成されたメタデータテーブルの例を図 30 に示す。図 30 のメタデータテーブルから、A200 という事物についての個人web ページのテキストに書かれた意見として、容量と用途という観点を取り上げられていること、用途としては海外出張用、国内出張用という2つの用途について評価されていること、容量は海外出張用には不十分（男性1名）と評価され、国内出張用にはあまりに大きい（女性1名）、十分だ（男性1名）と評価が分かれていることがわかる。

#### 【0069】

このように本実施の形態によれば、テキスト中の事物に関する記述内容を、意味的分類を用いて抽出し、観点と記述として整理した上で、さらに、要素メタデータやテキストの書誌情報であるソース情報およびテキストの作者に関するユーザ情報をも用いて客観性・信頼性の評価結果を示し、またユーザが要素メタデータに加えてソース情報から抽出したメタデータまたはユーザ情報から抽出したメタデータをも用いて必要な情報を指定することにより、記述された情報を精度良く抽出するだけでなく、抽出結果の個々の情報をユーザが適切に判断し、所望の情報を容易に選択できるようにすることができる。

#### 【産業上の利用可能性】

#### 【0070】

本発明にかかる情報抽出システムは、メタデータ抽出部、観点・記述抽出部、意味属性付与部、話題事物推定部、客観性・信頼性判定部、メタデータ照合部、メタデータ統合部、メタデータテーブル生成部を有し、情報抽出システム、情報検索システムとして有用である。また、情報分析/評価システム、情報配信システム等の用途にも応用できる。

#### 【図面の簡単な説明】

#### 【0071】

- 【図 1】本発明の実施の形態 1 に係る情報抽出システムの構成を示すブロック図
- 【図 2】実施の形態 1 に係る情報抽出システムにおける、テキストから要素メタデータを抽出するまでの一連の処理の流れを示す説明図
- 【図 3】実施の形態 1 に係る情報抽出システムにおける、観点・記述抽出規則と規則の構成要素定義の例を示す図
- 【図 4】実施の形態 1 に係る情報抽出システムにおける、統合メタデータの例を示す図
- 【図 5】本発明の実施の形態 2 に係る情報抽出システムの構成を示すブロック図
- 【図 6】実施の形態 2 に係る情報抽出システムにおける、入力されたテキストと意味属性を付与されたテキストの例を示す図
- 【図 7】実施の形態 2 に係る情報抽出システムにおける、意味属性付与規則の例と意味属性付与規則構成要素定義の例を示す図
- 【図 8】実施の形態 2 に係る情報抽出システムにおける、意味属性付きテキストの例と観点・記述認定例を示す図
- 【図 9】実施の形態 2 に係る情報抽出システムにおける、観点・記述の抽出のための規則の例を示す図
- 【図 10】実施の形態 2 に係る情報抽出システムにおける、要素メタデータ抽出結果の例を示す図
- 【図 11】実施の形態 2 に係る情報抽出システムにおける、統合メタデータの例を示す図
- 【図 12】本発明の実施の形態 3 に係る情報抽出システムの構成を示すブロック図
- 【図 13】実施の形態 3 に係る情報抽出システムにおける、観点・記述の認定結果と要素メタデータの抽出結果を示す図
- 【図 14】実施の形態 3 に係る情報抽出システムにおける、話題事物推定規則と話題事物推定規則構成要素定義の例を示す図
- 【図 15】実施の形態 3 に係る情報抽出システムにおける、推定した話題事物の例を示す図
- 【図 16】実施の形態 3 に係る情報抽出システムにおける、統合メタデータの例を示す図
- 【図 17】実施の形態 3 に係る情報抽出システムにおける、メタデータテーブルの例を示す図
- 【図 18】本発明の実施の形態 4 に係る情報抽出システムの構成を示すブロック図
- 【図 19】実施の形態 4 に係る情報抽出システムにおける、入力データ判定規則の例を示す図
- 【図 20】実施の形態 4 に係る情報抽出システムにおける、テキストのソース情報、ユーザ情報の例と意味属性付きソース情報、意味属性付きユーザ情報の例を示す図
- 【図 21】実施の形態 4 に係る情報抽出システムにおける、ソース情報意味属性付与規則、ユーザ意味属性付与規則の例を示す図
- 【図 22】実施の形態 4 に係る情報抽出システムにおける、ソース観点・記述抽出規則、ユーザ観点・記述抽出規則の例を示す図
- 【図 23】実施の形態 4 に係る情報抽出システムにおける、ソースメタデータ抽出結果、ユーザメタデータ抽出結果の例を示す図
- 【図 24】実施の形態 4 に係る情報抽出システムにおける、客観性・信頼性判定規則と客観性・信頼性判定規則構成要素定義の例を示す図
- 【図 25】実施の形態 4 に係る情報抽出システムにおける、テキストの例と意味属性付きテキストの例を示す図
- 【図 26】実施の形態 4 に係る情報抽出システムにおける、観点・記述抽出規則例と観点・記述構成要素定義例を示す図
- 【図 27】実施の形態 4 に係る情報抽出システムにおける、要素メタデータ抽出結果の例を示す図

【図 2 8】実施の形態 4 に係る情報抽出システムにおける、客観性・信頼性判定結果の例を示す図

【図 2 9】実施の形態 4 に係る情報抽出システムにおける、メタデータ統合結果の例を示す図

【図 3 0】実施の形態 4 に係る情報抽出システムにおける、メタデータテーブルの例を示す図

【符号の説明】

【 0 0 7 2 】

1 0 0 , 2 0 0 , 3 0 0 , 4 0 0    情報抽出システム

1 0 2    入力部

1 0 4    メタデータ抽出部

1 0 6    メタデータ照合部

1 0 8    メタデータ統合部

1 1 0    統合メタデータ記憶部

1 2 0    観点・記述抽出部

1 2 2    観点・記述抽出規則記憶部

2 0 2    意味属性付与部

2 0 4    意味属性付与規則記憶部

2 0 6    意味属性付きテキスト記憶部

3 0 2    ユーザ要求処理部

3 0 4    メタデータテーブル生成部

3 0 6    メタデータテーブル出力部

3 1 0    話題事物推定部

3 1 2    話題事物推定規則記憶部

4 0 2    入力データ判定部

4 0 4    入力データ判定規則記憶部

4 0 6    ソース意味属性付与規則記憶部

4 0 8    意味属性付きソース情報記憶部

4 1 0    ソース観点・記述抽出規則記憶部

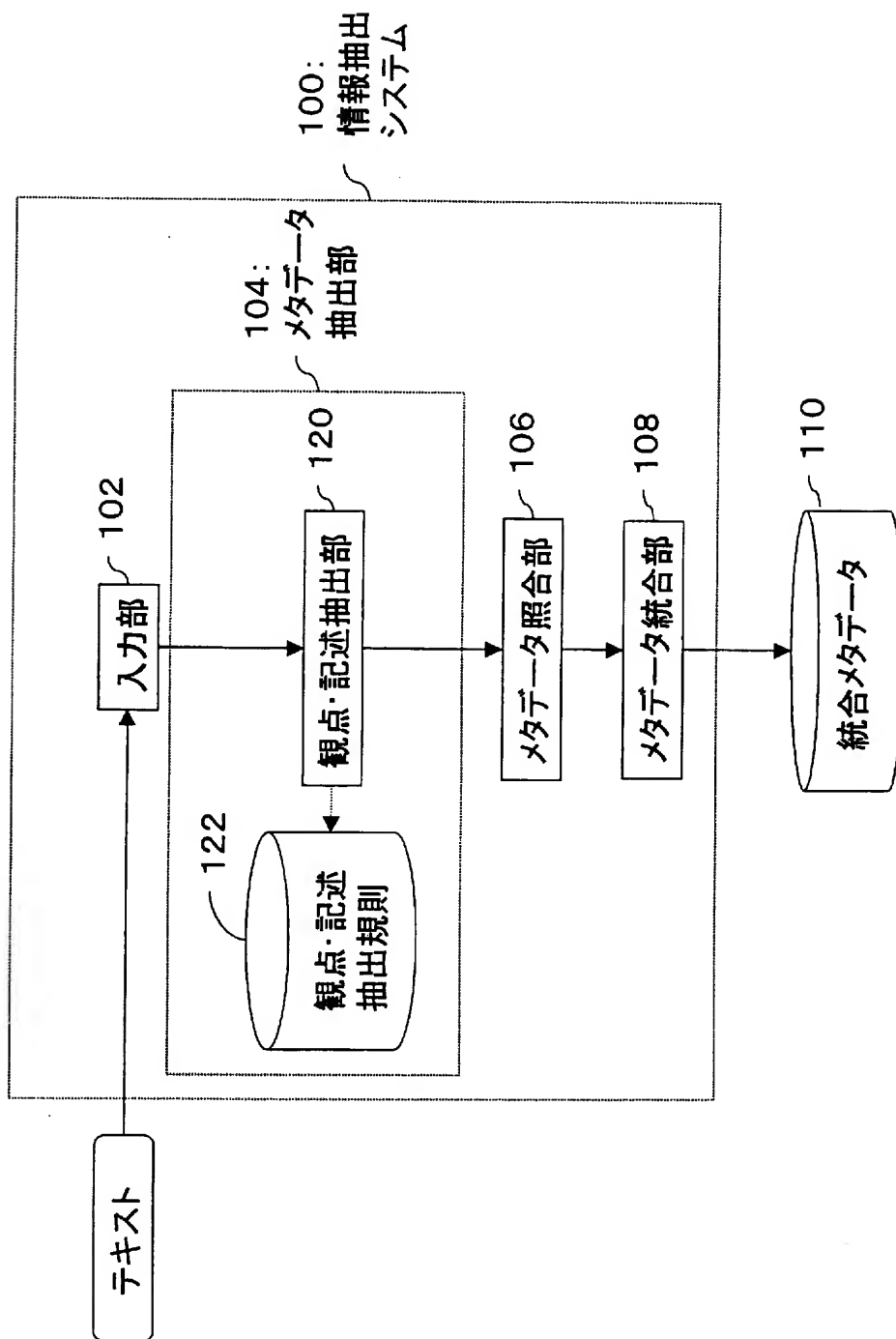
4 1 2    客観性・信頼性判定部

4 1 4    客観性・信頼性判定規則記憶部

4 1 6    ユーザ意味属性付与規則記憶部

4 1 8    意味属性付きユーザ情報記憶部

4 2 0    ユーザ観点・記述抽出規則記憶部



テキスト1

A社の小型バッグA20は開口部がかなり大きい。ファスナーの開閉も滑らかだ。皮の感触はしっとりやさしい。

テキスト2

A20は皮の手触りが驚くほどなめらかだ。色合いもしっとりやさしい。



観点・記述  
認定例

A社の小型バッグA20は<VIEW1>開口部</VIEW1>が<DESC1>かなり大きい</DESC1>。<VIEW2>ファスナーの開閉</VIEW2>も<DESC2>滑らかだ</DESC2>。<VIEW3>皮の感触</VIEW3>は<DESC3>しっとりやさしい</DESC3>。

A20は<VIEW1>皮の手触り</VIEW1>が<DESC1>驚くほどなめらかだ</DESC1>。<VIEW2>色合い</VIEW2>も<DESC2>しっとりやさしい</DESC2>。



要素メタデータ抽出結果

観点	記述	要素メタデータID
開口部	かなり大きい	1-1
ファスナーの開閉	滑らかだ	1-2
皮の感触	しっとりやさしい	1-3
皮の手触り	驚くほどなめらかだ	2-1
色合い	しっとりやさしい	2-2

観点・記述  
抽出規則例

規則	パターン	観点	記述
1	は({漢字/平仮名連続})[かも]({漢字/平仮名連続}){形容詞語尾1})	\$1	\$2
2	は({漢字/平仮名連続})[かも]({漢字/平仮名連続}){形容動詞語尾1})	\$1	\$2

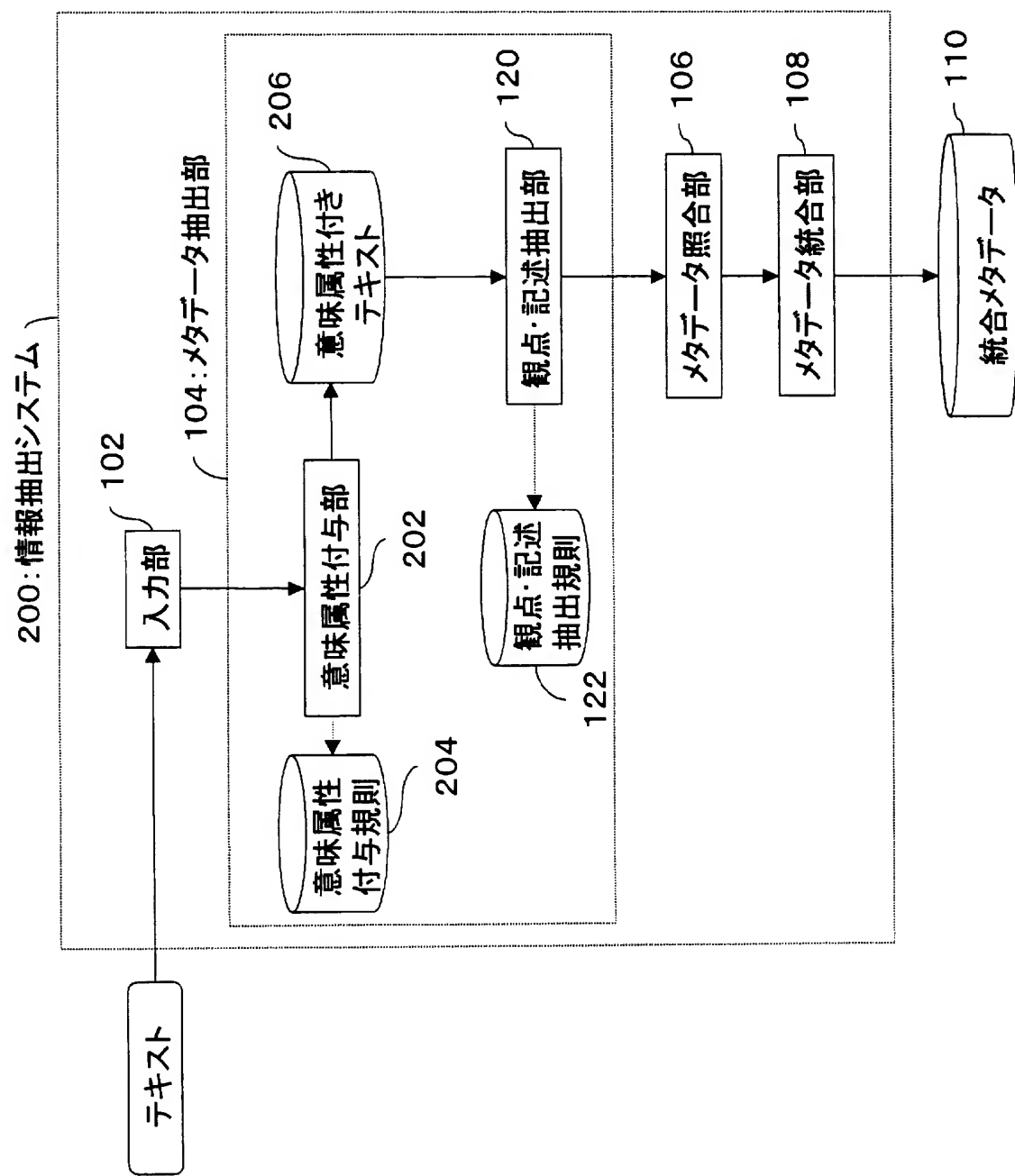
観点・記述抽出規則構成要素の定義例

- 形容動詞語尾1: だろ, だ, で, だっ, に, な
- 形容詞語尾1: く, かっ, う, ゆう, い
- [ ]: [ ] 内のいずれか
- ( ): 後方参照(順に"\$ {整数}"で参照される)
- \$ {整数}: 変数(パターンで {整数} 番目に"(") でくられた部分にマッチする文字列)

## メタデータ統合結果

観点	記述	観点・記述対ID
開口部	かなり大きい	1－1
ファスナーの開閉	滑らか	1－2
皮の感触	しっとりやさしい	1－3
	驚くほどなめらか	2－1
色合い	しっとりやさしい	2－2





テキスト例

テキスト1	A社のバッグA200は容量が20リットルと大きい。
テキスト2	バッグA200は容量が不十分だと思う。



意味属性付与例

テキスト1	<ORGANIZATION type=company>A社</ORGANIZATION><PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200</PROD_NAME>は<QUANT_TYPE>容量</QUANT_TYPE>が<QUANT unit=1, val=20>20リットル</QUANT>と大きい。
テキスト2	<PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200</PROD_NAME>は<QUANT_TYPE>容量</QUANT_TYPE>が不十分だと思う。

意味属性付与規則例

規則	パターン	対象部分	意味属性	詳細情報
1	{{数字連続}} ({{数量単位}})	\$1 \$2	QUANT	・unit: \$2またはその正規化形 ・val: \$1の数値表現を正規化した値
2	{{数量分類}}	\$1	QUANT_TYPE	
3	{{英字連続}} {会社名後接表現})	\$1	ORGANIZATION	・type:company
4	{{製品分類名}}	\$1	PROD_TYPE	
5	{{製品分類名}} {英数字連続})	\$1	PROD_NAME	

意味属性付与規則構成要素の定義例

- 会社名後接表現: 社
- 数量分類: 容量
- 数量単位: リットル、メートル、グラム、…
- 製品分類名: バッグ、シューズ、帽子、…

意味属性付きテキスト例

テキスト1	<ORGANIZATION type=company>A社</ORGANIZATION>の<PROD_TYPE> バッグ</PROD_TYPE><PROD_NAME>A200</PROD_NAME>は <QUANT_TYPE>容量</QUANT_TYPE>が<QUANT unit=l, val=20>20リットル </QUANT>と大きい。
テキスト2	<PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200</PROD_NAME> は<QUANT_TYPE>容量</QUANT_TYPE>が不十分だと思う。



観点・記述認定例

テキスト1	<DESC1><ORGANIZATION type=company>A社</ORGANIZATION></DESC1> の<DESC2><PROD_TYPE>バッグ</PROD_TYPE></DESC2> <DESC3> <PROD_NAME>A200</PROD_NAME></DESC3><VIEW4><QUANT_TYPE> 容量</QUANT_TYPE></VIEW4>が<DESC4a><QUANT unit=l, val=20>20リット ル</QUANT></DESC4a>と<DESC4b>大きい</DESC4b>。
テキスト2	<DESC1><PROD_TYPE>バッグ</PROD_TYPE></DESC1><DESC2> <PROD_NAME>A200</PROD_NAME></DESC2><VIEW3><QUANT_TYPE> 容量</QUANT_TYPE></VIEW3>が<DESC3>不十分だ</DESC3>と思う。

観点・記述抽出規則例

規則	パターン	観点	記述	注意
1	<QUANT_TYPE>({タグ開始記号以外の任意文字列})</QUANT_TYPE>[がは]<QUANT>({タグ開始記号以外の任意文字列})</QUANT>と({漢字/平仮名連続}){形容詞語尾1})	\$1	\$2,\$3	
2	<QUANT_TYPE>({タグ開始記号以外の任意文字列})</QUANT_TYPE>[がは]({漢字/平仮名連続}){形容動詞語尾1})	\$1	\$2	
3	{観点開始タグ・記述開始タグ以外の任意文字列}*<({意味属性})>({タグ開始記号以外の任意文字列})<({意味属性})>	\$1の意味属性の別名	\$2	※1 ※2

※1:他の規則の適用後に適用する。  
※2:意味属性は開始タグのみ詳細情報を含む。

観点・記述抽出規則構成要素定義例

+:直前のパターン要素の1回以上の繰り返し

\*:直前のパターン要素の0回以上の繰り返し

観点開始タグ:<VIEW[0-9a-z]++>

記述開始タグ:<DESC[0-9a-z]++>

タグ開始記号:"<"

形容動詞語尾1:だろ,だ,で,だっ,に,な

形容詞語尾1:く,かつ,う,ゆう,い

意味属性:

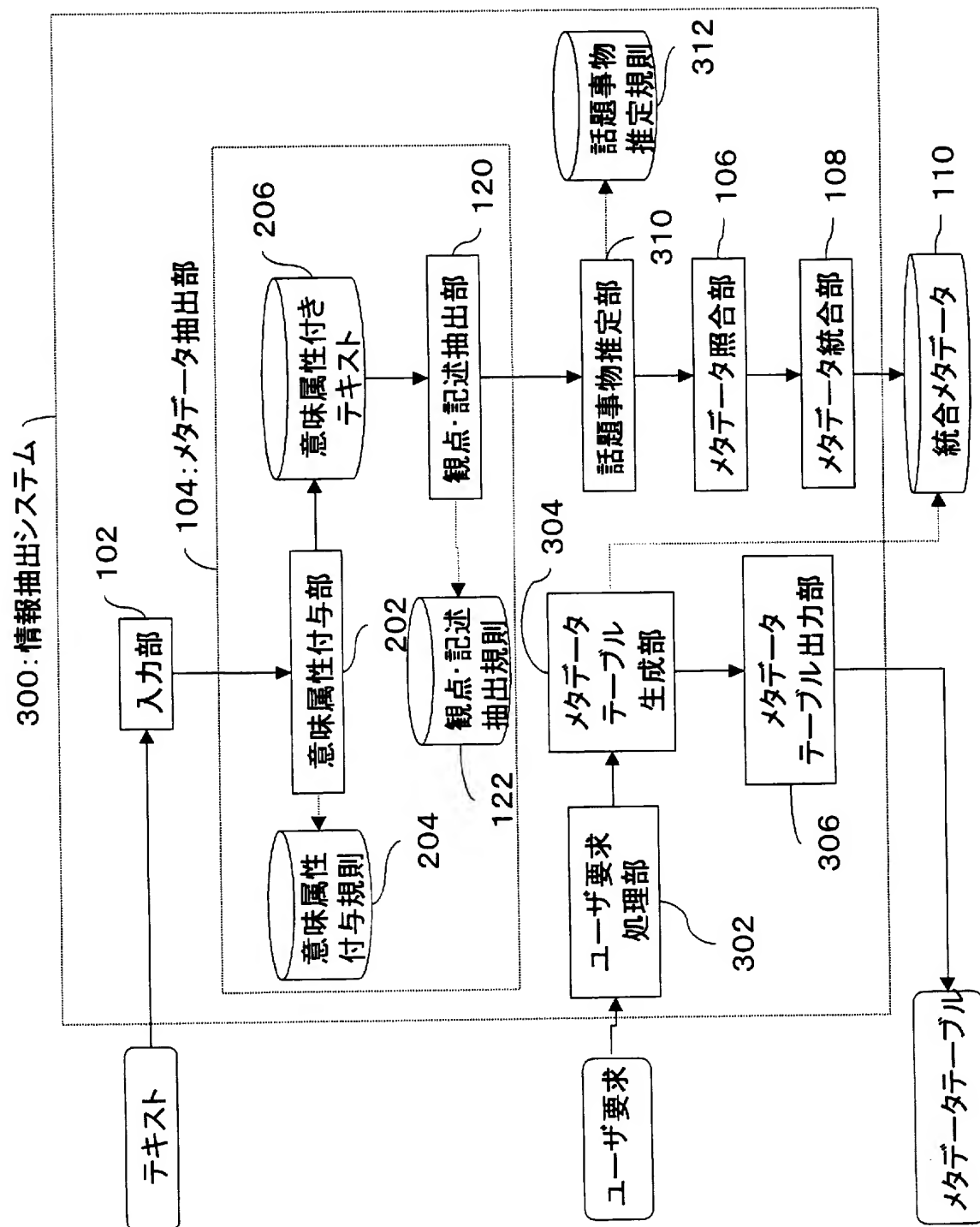
ORGANIZATION,ORGANIZATION type=company,  
QUANT\_TYPE,QUANT,PROD\_TYPE,PROD\_NAME,PERSON,DATE,TIME,PERIOD,...  
意味属性の別名:  
ORGANIZATION:組織名  
ORGANIZATION type=company:会社名  
QUANT:数量  
PROD\_TYPE:製品分類  
...

要素メタデータ抽出結果

観点	記述	意味属性	詳細情報	要素メタ データID
会社名	A社	ORGANIZATION	type=company	1-1
製品分類	バッグ	PROD_TYPE		1-2
製品名	A200	PROD_NAME		1-3
容量		QUANT_TYPE		-
	20リットル	QUANT	unit=l, val=20	1-4a
	大きい			1-4b
製品分類	バッグ	PROD_TYPE		2-1
製品名	A200	PROD_NAME		2-2
容量		QUANT_TYPE		2-3
	不十分だ			

メタデータ統合結果

観点	記述	意味属性	詳細情報	要素メタ データID
会社名	A社	ORGANIZATION	type=company	1-1
製品分類	バッグ	PROD_TYPE		1-2
				2-1
製品名	A200	PROD_NAME		1-3
				2-2
容量		QUANT_TYPE		-
		QUANT	unit=l, val=20	1-4a
		大きい		1-4b
		不十分だ		2-3





## 観点・記述認定例

## テキスト1

<DESC1><PROD\_TYPE>バッグ</PROD\_TYPE></DESC1><DESC2><PROD\_NAME>A200  
 </PROD\_NAME></DESC2>は<VIEW3><QUANT\_TYPE>容量</QUANT\_TYPE></VIEW3>が  
 <DESC3>不十分だ</DESC3>し、<DESC4><PROD\_TYPE>バッグ</PROD\_TYPE></DESC4>  
 <DESC5><PROD\_NAME>A300</PROD\_NAME></DESC5>は<VIEW6><QUANT\_TYPE>容量  
 </QUANT\_TYPE></VIEW6>が<DESC6>あまりに大きい</DESC6>。

## テキスト2

<DESC1><PROD\_TYPE>バッグ</PROD\_TYPE></DESC1><DESC2><PROD\_NAME>A200  
 </PROD\_NAME></DESC2>は<VIEW3><QUANT\_TYPE>容量</QUANT\_TYPE></VIEW3>が  
 <DESC3><QUANT unit=l value=20>20リットル</QUANT></DESC3>で、<DESC4><PROD\_TYPE>  
 バッグ</PROD\_TYPE></DESC4><DESC5><PROD\_NAME>A300</PROD\_NAME></DESC5>の  
 <VIEW6><QUANT\_TYPE>容量</QUANT\_TYPE></VIEW6>は<DESC6>  
 <QUANT unit=l value=30>30リットル</QUANT></DESC6>。

## 要素メタデータ抽出結果例

観点	記述	意味属性	詳細情報	要素メタデータID
製品分類	バッグ	PROD_TYPE		1-1
製品名	A200	PROD_NAME		1-2
容量	不十分だ	QUANT_TYPE		1-3
製品分類	バッグ	PROD_TYPE		1-4
製品名	A300	PROD_NAME		1-5
容量	あまりに大きい	QUANT_TYPE		1-6
製品分類	バッグ	PROD_TYPE		2-1
製品名	A200	PROD_NAME		2-2
容量	20リットル	QUANT	unit=l, val=20	2-3
製品分類	バッグ	PROD_TYPE		2-4
製品名	A300	PROD_NAME		2-5
容量	30リットル	QUANT	unit=l, val=30	2-6

話題事物推定規則の例

規則	条件	推定話題事物
1	<DESC[0-9]+><{タグ構成文字列}>{話題事物候補}<{タグ構成文字列}></DESC[0-9]+>	記述\$1の要素メタデータの話題事物は\$1
2	{(話題事物候補)}<{タグ構成文字列}></DESC[0-9]+>[の は]<(VIEW[0-9]+)><{タグ構成文字列}><{任意文字列}><{任意文字列}><{タグ構成文字列}><(VIEW[0-9]+)>[は が]<(DESC[0-9]+)>({任意文字列})</DESC[0-9]+>	観点\$3、記述\$6の要素メタデータの話題事物は\$1
3	ただし、\$2と\$4,\$5と\$7の文字列は同一であること  <(DESC[0-9]+)><{タグ構成文字列}>({任意文字列})<{タグ構成文字列}></DESC[0-9]+><DESC[0-9]+><{タグ構成文字列}>({話題事物候補})</{タグ構成文字列}>	記述\$2の要素メタデータの話題事物は\$4
	ただし、\$1と\$3の文字列は同じであること	

話題事物推定規則構成要素定義の例

+ : 直前のパターン要素の1回以上の繰り返し  
タグ構成文字列 : [半角英数字,スペース,""]+

話題事物推定例

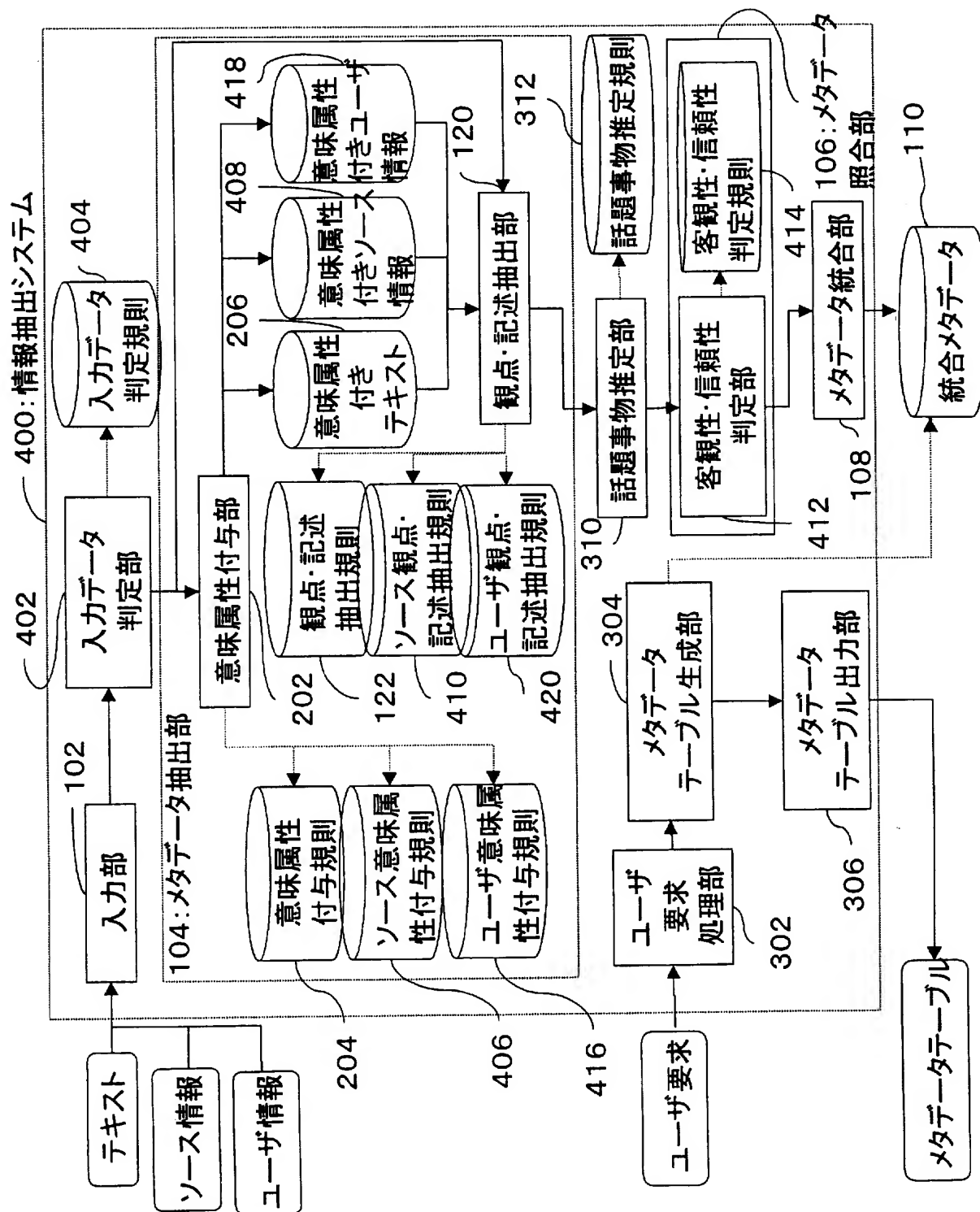
観点	記述	意味属性	詳細情報	要素メタデータID	話題事物
製品分類	バッグ	PROD_TYPE		1-1	A200
製品名	A200	PROD_NAME		1-2	A200
容量	不十分だ	QUANT_TYPE		1-3	A200
製品分類	バッグ	PROD_TYPE		1-4	A300
製品名	A300	PROD_NAME		1-5	A300
容量	あまりに大きい	QUANT_TYPE		1-6	A300
製品分類	バッグ	PROD_TYPE		2-1	A200
製品名	A200	PROD_NAME		2-2	A200
容量	20リットル	QUANT_TYPE		2-3	A200
		QUANT	unit=l, val=20		
製品分類	バッグ	PROD_TYPE		2-4	A300
製品名	A300	PROD_NAME		2-5	A300
容量	30リットル	QUANT_TYPE		2-6	A300
		QUANT	unit=l, val=30		

統合結果例

話題 事物	観点	記述	意味属性	詳細情報	要素メタデー タID
A200	製品分類	バッグ	PROD_TYPE		1-1 2-1
	製品名	A200	PROD_NAME		1-2 2-2
	容量	不十分だ	QUANT_TYPE		-
					1-3
		20リットル	QUANT	unit=l, val=20	2-3
A300	製品分類	バッグ	PROD_TYPE		1-4 2-4
	製品名	A300	PROD_NAME		1-5 2-5
	容量	あまりに大きい	QUANT_TYPE		-
					1-6
		30リットル	QUANT	unit=l, val=30	2-6

統合メタデータテーブルの例

話題 事物	観点	記述	意味属性	詳細情報	要素メタデー タID
A200	容量	不十分だ	QUANT_TYPE		-
					1-3
		20リットル	QUANT	unit=1, val=20	2-3



入力データ判定規則の例

規則	条件	判定内容
1	ファイル名 (.+).txt	テキスト識別子:\$1 データの種別:テキストのみ
2	ファイル名 (.+).bib	テキスト識別子:\$1 データの種別:ソース情報のみ
3	ファイル名 (.+).usr	テキスト識別子:\$1 データの種別:ユーザ情報のみ
4	ファイル名 (.+).msc	テキスト識別子:\$1 データの種別:テキスト、ソース情報、ユーザ情報の少なくとも 2種以上を含む
5	ファイルの第1行 bib:([0-9]+)-([0-9]+)	ソース情報の範囲:\$1行から\$2行
6	ファイルの第1行 usr:([0-9]+)-([0-9]+)	ユーザ情報の範囲:\$1行から\$2行
7	ファイルの第2行 no tagging:l	意味属性付与処理:不要

.:改行文字以外の任意の1文字

### ソース情報の例

#### ソース情報 1

<http://www.aaa.co.jp/article1>

作成日：2003年10月1日

#### ソース情報 2

<http://www.xxx.yyy.jp/~zzz>

作成日：2003年5月1日

### ユーザ情報の例

#### ユーザ情報 1

会社名：aaa

#### ユーザ情報 2

作者zzz。20代男性。

### 意味属性付きソース情報の例

#### ソース情報 1

<URL type=company><http://www.aaa.co.jp/article1></URL>

<DATE value=2003:10:01>2003年10月1日</DATE>

### 意味属性付きユーザ情報の例

#### ユーザ情報 2

作者<AUTHOR>zzz</AUTHOR>。<AGE value=20:29>20代

</AGE><GENDER type=M>男性</GENDER>。



ソース意味属性付与規則の例

規則	パターン	意味分類	詳細情報
1	(http://.*.co*.jp.*)	URL(webページ)	type=company
2	((数字4桁)年(数字1 ~2桁)月(数字1~2 桁))	DATE	value=\$1.\$2.\$3 ただし\$2,\$3が1桁の場合0 を前につける

¥,“,””(ピリオド)

ユーザ意味属性付与規則の例

規則	パターン	意味分類	詳細情報
1	作者[:]*([~。]+)	AUTHOR	
2	((数字1桁)0代	AGE	value=(\$1*10): (\$1*10+ 9)
3	男性	GENDER	type=M

[^...]:[]内の文字に一致しない文字

ソース観点・記述抽出規則の例

規則	パターン	観点	記述
1	<URL type=company>{タグ開始記号以外の任意文字列}</URL>	テキストの入手元	\$1
2	<URL type=company>{タグ開始記号以外の任意文字列}</URL>	テキストの種類	会社webページ(\$1)
3	(作成日):<DATE>{タグ開始記号以外の任意文字列}</DATE>	\$1	\$2

ユーザ観点・記述抽出規則の例

規則	パターン	観点	記述
1	<AUTHOR>{タグ開始記号以外の任意文字列}</AUTHOR>	作者	\$1
2	<AGE value=(20:29)>{タグ開始記号以外の任意文字列}</AGE>	年齢	\$1
3	<GENDER type=M>{タグ開始記号以外の任意文字列}</GENDER>	性別	男性

ソースメタデータ抽出結果の例

観点	記述	意味属性	詳細情報	ソースメタデータID
テキストの入手元	http://www.aaa.co.jp/article1	URL(webページ)	type=company	1-S1
テキストの分類	会社webページ (http://www.aaa.co.jp/article1)	URL(webページ)	type=company	1-S2
作成日	2003年10月1日	DATE	value=2003:10:01	1-S3

ユーザメタデータ抽出結果の例

観点	記述	意味属性	詳細情報	ユーザメタデータID
作者	zzz	AUTHOR		2-U1
年齢	20-29	AGE	value=20:29	2-U2
性別	男性	GENDER	type=M	2-U3

客観性・信頼性判定規則例

規 則	条 件					客 観 性	信 頼 性
	テキストの要素メタデータ		ソースメタデータ		その他		
	観点	記述の意味分類	観点	記述			
1	組織名					1	1
2	製品分類					1	1
3	製品名					1	1
4	用途	USAGE	テキストの分類	会社webページ/新聞記事		1	1
5	用途	USAGE	テキストの分類	個人webページ		0	0.5
6	容量	QUANT	テキストの分類	会社webページ/新聞記事		1	1
7	容量	QUANT	テキストの分類	個人webページ		1	0.9
8	容量	なし、またはQUANT以外			文末が{不確定表現1}	0	0.2
9	容量	なし、またはQUANT以外	テキストの分類	会社webページ/新聞記事	文末が{不確定表現1}以外	0	0.5
10	容量	なし、またはQUANT以外	テキストの分類	個人webページ	文末が{不確定表現1}以外	0	0.3

客観性・信頼性判定規則構成要素定義例

不確定表現1: と思う, と思える, と思われる, かもしれない, らしい

テキスト例

テキスト1	バッグA200は容量が20リットルと大きい。
テキスト2	バッグA200の容量は海外出張用には不十分だと思う。
テキスト3	バッグA200の容量は国内出張用にはあまりに大きい。
テキスト4	バッグA200の容量は国内出張用には十分だ。

意味属性付きテキスト例

テキスト1	<PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200</PROD_NAME>は<QUANT_TYPE>容量</QUANT_TYPE>が<QUANT unit=l, val=20>20リットル</QUANT>と大きい。
テキスト2	<PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200<PROD_NAME>の<QUANT_TYPE>容量</QUANT_TYPE>は<USAGE>海外出張用</USAGE>には不十分だと思う。
テキスト3	<PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200<PROD_NAME>の<QUANT_TYPE>容量</QUANT_TYPE>は<USAGE>国内出張用</USAGE>にはあまりに大きい。
テキスト4	<PROD_TYPE>バッグ</PROD_TYPE><PROD_NAME>A200<PROD_NAME>の<QUANT_TYPE>容量</QUANT_TYPE>は<USAGE>国内出張用</USAGE>には十分だ。

観点・記述抽出規則例

規則	パターン	観点	記述	注意
1	<QUANT_TYPE>({タグ開始記号以外の任意文字列})(<QUANT_TYPE>[がは]<QUANT>({タグ開始記号以外の任意文字列})<QUANT>と({漢字/平仮名連続}{形容詞語尾1}))	\$1	\$2  \$3	
2	<QUANT_TYPE>({タグ開始記号以外の任意文字列})(<QUANT_TYPE>[がは]<USAGE>({タグ開始記号以外の任意文字列}用)<USAGE>(に として)[はも]({漢字/平仮名連続}{形容詞語尾1}))	\$1	\$2&&\$3	
3	<QUANT_TYPE>({タグ開始記号以外の任意文字列})<QUANT_TYPE>[がは]({漢字/平仮名連続}{形容動詞語尾1}))	\$1	\$2	
4	{観点開始タグ・記述開始タグ以外の任意文字列}*(<({意味属性})>({タグ開始記号以外の任意文字列})<({意味属性})>	\$1の意味属性の別名	\$2	※1 ※2

※1: 他の規則の適用後に適用する。  
※2: 意味属性は開始タグのみ詳細情報を含む。

観点・記述抽出規則構成要素定義例

||: 同じ観点についての複数の異なる記述  
&&: 同じ観点についての関連する記述

要素メタデータ抽出結果例

観点	記述	意味属性	詳細情報	要素メタデータID	話題事物
製品分類	バッグ	PROD_TYPE		1-1	A-200
製品名	A200	PROD_NAME		1-2	A-200
容量	20リットル 大きい	QUANT_TYPE		-	A-200
		QUANT	unit=l, val=20	1-3a	
		-		1-3b	
製品分類	バッグ	PROD_TYPE		2-1	A-200
製品名	A200	PROD_NAME		2-2	A-200
容量	海外出張用 不十分だ	QUANT_TYPE		2-3	A-200
		USAGE			
		-			
用途	海外出張用	USAGE		2-4	A-200
製品分類	バッグ	PROD_TYPE		3-1	A-200
製品名	A200	PROD_NAME		3-2	A-200
容量	国内出張用 あまりに大きい	QUANT_TYPE		3-3	A-200
		USAGE			
		-			
用途	国内出張用	USAGE		3-4	A-200
製品分類	バッグ	PROD_TYPE		4-1	A-200
製品名	A200	PROD_NAME		4-2	A-200
容量	国内出張用 十分だ	QUANT_TYPE		4-3	A-200
		USAGE			
		-			
用途	国内出張用	USAGE		4-4	A-200

客観性・信頼性判定結果例

要素メタデータ							ソースメタデータ			ユーザメタデータ		
観点	記述	意味属性	要素メタデータID	客観性	信頼性	観点	記述	ソースメタデータID	観点	記述	ユーザメタデータID	
製品分類	バッグ	PROD_TYPE	1-1	1.0	1.0	テキストの分類	会社webページ	1-S2	性別	-	-	
	A200	PROD_NAME	1-2	1.0	1.0							
	容量		QUANT_TYPE	-	-		-					
		20リットル	QUANT	1-3a	1.0		1.0					
		大きい	-	1-3b	0		0.5					
製品分類	バッグ	PROD_TYPE	2-1	1.0	1.0		個人webページ	2-S2	2-U3	男性		
	A200	PROD_NAME	2-2	1.0	1.0							
	容量		QUANT_TYPE	-	-							-
		海外出張用	USAGE	2-3	0							0.2
		不十分だ	-	-	-							-
用途	海外出張用	USAGE	2-4	0	0.5	個人webページ	3-S2	3-U3	女性			
	バッグ	PROD_TYPE	3-1	1.0	1.0							
	A200	PROD_NAME	3-2	1.0	1.0							
	容量		QUANT_TYPE	-	-						-	
		国内出張用	USAGE	3-3	0						0.3	
あまりに大きい		-	-	-	-							
用途	国内出張用	USAGE	3-4	0	0.5	個人webページ	4-S2	4-U3	男性			
	バッグ	PROD_TYPE	4-1	1.0	1.0							
	A200	PROD_NAME	4-2	1.0	1.0							
	容量		QUANT_TYPE	-	-						-	
		国内出張用	USAGE	4-3	0						0.3	
十分だ		-	-	-	-							
用途	国内出張用	USAGE	4-4	0	0.5							



メタデータ統合結果例

要素メタデータ										ソースメタデータ			ユーザメタデータ		
話題 事物	観 点	記 述	意味属性	要素メタデー タID	客観性	信頼性	観 点	記 述	ソースメタ データID	観 点	記 述	ユーザメタ データID			
A200	製 品 分 類	バッグ	PROD_TYPE	1-1	1.0	1.0	テ キ ス ト の 分 類	会社webペー ジ	1-S2	性別	-	-			
				2-1	個人webペー ジ	2-S2		男性	2-U3						
				3-1		3-S2		女性	3-U3						
				4-1		4-S2		男性	4-U3						
	製 品 名	A200	PROD_NAME	1-2	1.0	1.0		会社webペー ジ	1-S2		-	-			
				2-2	個人webペー ジ	2-S2		男性	2-U3						
				3-2		3-S2		女性	3-U3						
				4-2		4-S2		男性	4-U3						
	容 量			QUANT_TYPE	-								-	-	
					20リットル	QUANT		1-3a	1.0		1.0	会社webペー ジ	1-S2	-	-
					大きい	-		1-3b	0		0.5	個人webペー ジ	2-S2	男性	2-U3
					海外出張用	USAGE		2-3	0		0.2		女性	3-U3	
					不十分だ	-							男性	4-U3	
					国内出張用	USAGE		3-3	0		0.3	個人webペー ジ	3-S2	男性	2-U3
					あまりに大きい	-							女性	3-U3	
					国内出張用	USAGE		4-3	0		0.3		男性	4-U3	
					十分だ	-									
					海外出張用	USAGE		2-4	0		0.5	個人webペー ジ	2-S2	男性	2-U3
					国内出張用	USAGE		3-4					女性	3-U3	
				4-4			4-S2	男性	4-U3						

メタデータテーブル

要素メタデータ							ソースメタデータ			ユーザメタデータ		
話題 事物	親 点	記 述	意味属性	要素メタ データID	客観 性	信 類 性	親 点	記 述	ソースメタデー タID	親 点	記 述	ユーザメタ データID
A200	容 量		QUANT_TYPE	-	-	-	テキストの種類	個人ウェブページ		性別	-	
		海外出張用	USAGE	2-3	0	0.2			2-S2		男性	2-U3
		不十分だ	-									
		国内出張用	USAGE	3-3	0	0.3			3-S2		女性	3-U3
		あまりに大きい	-									
		国内出張用	USAGE	4-3	0	0.3			4-S2		男性	4-U3
		十分だ	-									
	用 途	海外出張用	USAGE	2-4	0	0.5			2-S2	男性	2-U3	
		国内出張用	USAGE	3-4					3-S2	女性	3-U3	
				4-4					4-S2	男性	4-U3	

【書類名】 要約書

【要約】

【課題】 テキスト中の事物に関する事実や意見の情報を精度良く、かつ、整理して抽出すること。

【解決手段】 テキスト中の事物に関する情報に意味属性を付与する意味属性付与手段と、観点と記述として整理して抽出する観点・記述抽出手段を有するメタデータ抽出手段、メタデータ間の類似度の比較を行うメタデータ照合手段、話題事物を推定するための話題事物推定手段、メタデータの照合結果に応じてメタデータを統合するメタデータ統合手段、メタデータの客観性・信頼性をテキストの書誌情報や作者に関する情報などを用いて判定する客観性・信頼性判定手段、を有する構成とする。これにより、テキスト中の事物に関する事実や意見の記述を精度良く抽出し、観点・記述として整理し、内容の対応関係や類似度や客観性・信頼性を詳細に比較・分析することができるようにし、かつ内容をわかりやすく提示することができる。

【選択図】 図 1

## 出願人履歴

0 0 0 0 0 5 8 2 1

19900828

新規登録

大阪府門真市大字門真 1 0 0 6 番地

松下電器産業株式会社